# Speech Command Recognition for NAO Robot


**Project Supervisor: Prof. G. C. Nandi**

**(Chairman, Robotics and Machine Intelligence Laboratory)**

**IEC2016012 - Nikhil Mundra**

**IEC2016026 - Agam Dwivedi**

**IEC2016027 - Bhanu Bhandari**

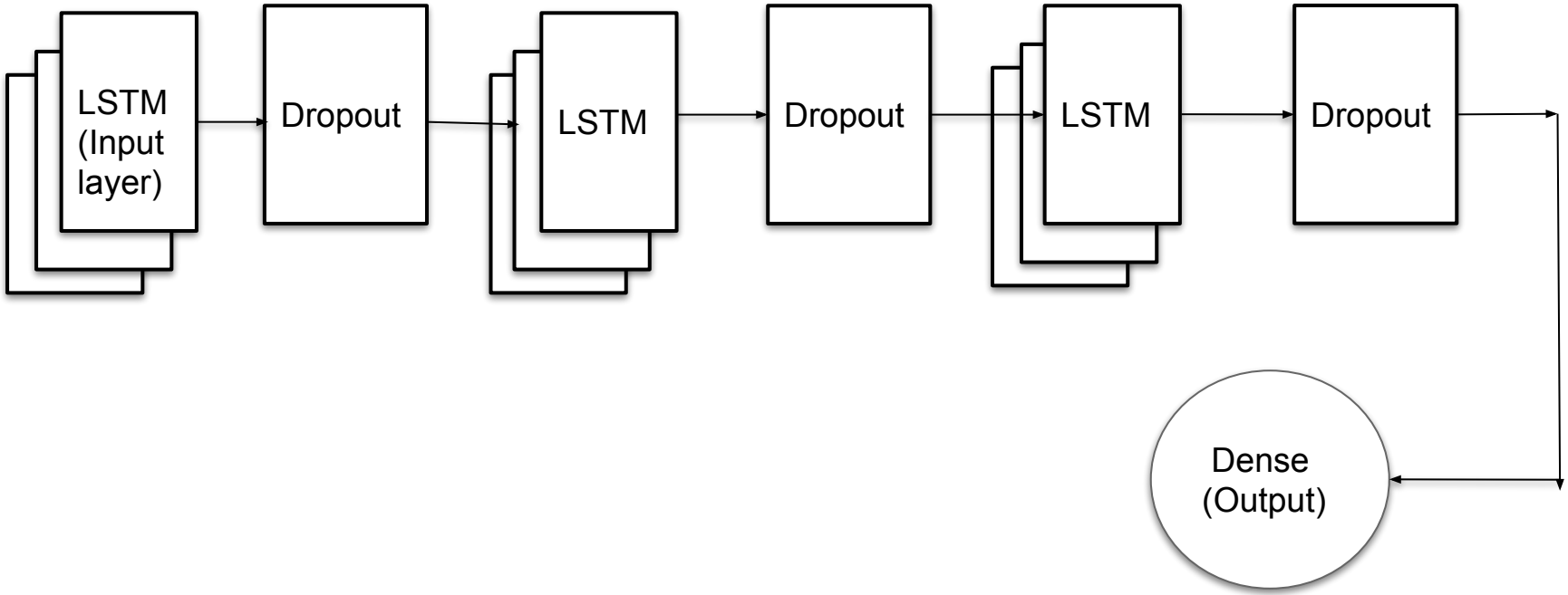
**IEC2016072 - Ruchin Agrawal**


# Problem statement

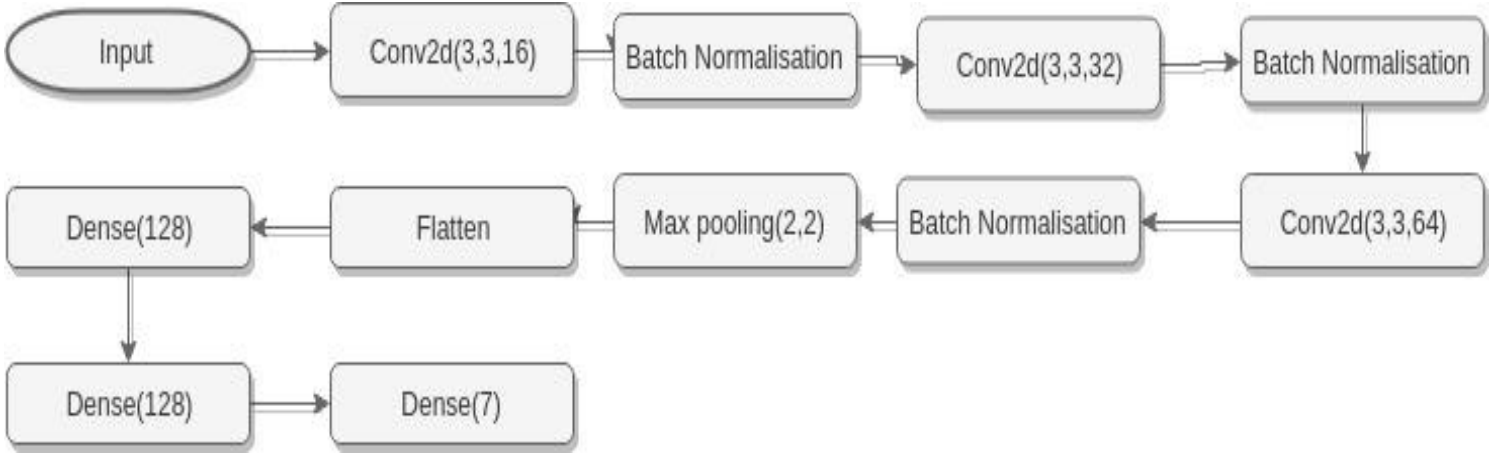
- To create Neural Network based architecture which allows understanding of simple speech commands, which trigger actions on the NAO Robot

- We have explored CNNs and RNNs in order to achieve high real-time accuracy, on both Hindi and English datasets

# LSTM Architecture



# CNN Architecture



# Dataset

**English**

- Tensorflow Speech Commands dataset
- **10 important action words**
- 18 auxiliary words
- 1700 varied utterances per word
- A number of accents, male/female voices, amplitudes, background noises

**Hindi**

- **Self created dataset**
- 10 utterances of 26 words with 10 speakers.
- **Many kinds of augmentations**
- 2600 iterations
- 6 male and 4 female voices

# Results

| | |
|---|---|
| **CNN with English Dataset**<br><br>Accuracy: 90.19% | **CNN with Hindi Dataset**<br><br>Accuracy: 79.94% |
| **LSTM with English Dataset**<br><br>Accuracy: 79.99% | **LSTM with Hindi Dataset**<br><br>Accuracy: Unsatisfactory |

# Analysis

- **Disadvantages of using Dropout**
- **Batch normalization vs Dropout**
- **Predicted results vs Real-time accuracy**
- **LSTM vs CNN**
- **Accuracy with different optimizers, activation functions, number of dense layers, kernel size**